

Change Detection in Streaming Data

Scientific Data Management Research Group, Lawrence Berkeley National Laboratory

01/20/2016

1 Project description

Advanced monitoring instruments are collecting a vast amount of data on electric power grid. Similarly, large science experiments and simulations are also generating huge quantities of data. Extracting features from such data streams is a critical need in understanding important phenomena including shock propagation in electrical power grid and detecting transient stars in astronomic observations. Based on our earlier experiences detecting changes in time series, we plan to explore techniques of identifying changes in live data streams.

2 Task Goals and requirements

During the summer of 2016, we plan to study the existing change detection literature and then explore the option of high-performance implementation of change detection techniques. The intern is expected to study the key implementation strategies and design choices. The intern will also be responsible for laying out the outline of implementation plan, and implement the key building blocks. This work will also involve interacting with staff and outside consultants to understand the impact of the design choices and demonstrate the pros and cons of the choices made. Task requirements are following:

- Strong analytical skills
- Proficient in programming languages C/C++
- Familiar with parallelization technology such as posix threading, openCL and MPI
- Good problem solving skills and communication skills

Students will learn about advanced statistics and data analysis techniques, debugging C/C++ applications, and help solving computational finance problems.

About the group

The Scientific Data Management (SDM) group develops technologies and tools for efficient data access and storage management of massive scientific data sets. We are currently developing storage resource management tools, data querying technologies, in situ feature extraction algorithms, along with software platforms for exascale data. The group also works closely with application scientists to address their data processing challenges. These tools and application development activities are backed by active research efforts on novel algorithms for emerging hardware platforms.